

## Research Article

# Multimodal Affective Computing to Enhance the User Experience of Educational Software Applications

**Jose Maria Garcia-Garcia** <sup>1</sup>, **Víctor M. R. Penichet** <sup>1</sup>, **María Dolores Lozano** <sup>1</sup>,  
**Juan Enrique Garrido** <sup>2</sup> and **Effie Lai-Chong Law**<sup>3</sup>

<sup>1</sup>Research Institute of Informatics, University of Castilla-La Mancha, Albacete, Spain

<sup>2</sup>Escuela Politécnica Superior, University of Lleida, Lleida, Spain

<sup>3</sup>Department of Informatics, University of Leicester, Leicester, UK

Correspondence should be addressed to María Dolores Lozano; [maria.lozano@uclm.es](mailto:maria.lozano@uclm.es)

Received 19 January 2018; Revised 8 July 2018; Accepted 5 August 2018; Published 13 September 2018

Academic Editor: Salvatore Carta

Copyright © 2018 Jose Maria Garcia-Garcia et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Affective computing is becoming more and more important as it enables to extend the possibilities of computing technologies by incorporating emotions. In fact, the detection of users' emotions has become one of the most important aspects regarding Affective Computing. In this paper, we present an educational software application that incorporates affective computing by detecting the users' emotional states to adapt its behaviour to the emotions sensed. This way, we aim at increasing users' engagement to keep them motivated for longer periods of time, thus improving their learning progress. To prove this, the application has been assessed with real users. The performance of a set of users using the proposed system has been compared with a control group that used the same system without implementing emotion detection. The outcomes of this evaluation have shown that our proposed system, incorporating affective computing, produced better results than the one used by the control group.

## 1. Introduction

In 1997, Rosalind W. Picard [1] defined Affective Computing as “computing that relates to, arises from, or influences emotions or other affective phenomena.” Since then, a general concern about the consideration of the emotional states of users for different purposes has arisen in different research fields (psychology [2, 3], marketing, computing, etc.).

Concretely, the underlying idea of Affective Computing is that computers that interact with humans need the ability to at least recognize affect [4]. Indeed, affective computing is a new field, with recent results in areas such as learning [5], information retrieval, communications [6], entertainment, design, health, marketing, decision-making, and human interaction where affective computing may be applied [7]. Different studies have proved the influence of emotions in consumers' behaviour [8] and decision-making activities [9].

In computer science research, we could study emotions from different perspectives. Picard mentioned that if we

want computers to be genuinely intelligent and to interact naturally with us, we must give computers the ability to recognize, understand, even to have and express emotions. In another different research work, Rosalind pointed out some inspiring challenges [10]: sensing and recognition, modelling, expression, ethics, and utility of considering affect in HCI. Studying such challenges still makes sense since there are gaps to be explored behind them. In human-computer interaction, emotion helps regulate and bias processes in a helpful way.

In this paper, we focus our research in the use of emotions to dynamically modify the behaviour of an educational software application according to the user feelings, as described in Section 3. This way, if the user is tired or stressed, the application will decrease its pace and, in some cases, the level of difficulty. On the other hand, if the user is getting bored, the application will increase the pace and the difficulty level so as to motivate the user to continue using the application.

Finally, we have assessed the application to prove that including emotion detection in the implementation of educational software applications considerably improves users' performance.

The rest of the paper is organized in the following sections: In Section 2, some background concepts and related works are presented. In Section 3, we describe the educational software application we have developed enhanced with affective computing-related technologies. Section 4 shows the evaluation process carried out to prove the benefits of the system developed. Finally, Section 5 presents some conclusions and final remarks.

## 2. Background Concepts and Related Works

In this section, a summary of the background concepts of affective computing and related technologies is put forward. We provide a comparison among the different ways of detecting emotions together with the technologies developed in this field.

**2.1. Affective Computing.** Rosalind Picard used the term "affective computing" for the first time in 1995 [11]. This technical report established the first ideas on this field. The aim was not to answer questions such as "what are emotions?," "what causes them?," or "why do we have them?," but to provide a definition of some terms in the field of affective computing.

As stated before, the term "affective computing" was finally set in 1997 as "computing that relates to, arises from, or deliberately influences emotion or other affective phenomena" [1]. More recently, we can find the definition of Affective computing as the study and development of systems and devices that can recognize, interpret, process, and simulate human affects [4]. In other words, any form of computing that has something to do with emotions. Due to the strong relation with emotions, their correct detection is the cornerstone of Affective Computing. Even though each type of technology works in a specific way, all of them share a common core in the way they work, since an emotion detector is, fundamentally, an automatic classifier.

The creation of an automatic classifier involves collecting information, extracting the features which are important for our purpose, and finally training the model so it can recognize and classify certain patterns [12]. Later, we can use the model to classify new data. For example, if we want to build a model to extract emotions of happiness and sadness from facial expressions, we have to feed the model with pictures of people smiling, tagged with "happiness" and pictures of people frowning, tagged with "sadness." After that, when it receives a picture of a person smiling, it identifies the shown emotion as "happiness," while pictures of people frowning will return "sadness" as a result.

Humans express their feelings through several channels: facial expressions, voices, body gestures and movements, and so on. Even our bodies experiment visible physical reactions to emotions (breath and heart rate, pupil's size, etc.).

Because of the high potential of knowing how the user is feeling, this kind of technology (emotion detection) has experienced an outburst in the business sector. Many technology companies have recently emerged, focused exclusively on developing technologies capable of detecting emotions from specific input. In the following sections, we present a brief review of each kind of affective information channel, along with some existing technologies capable of detecting this kind of information.

**2.2. Emotion Detection Technologies.** This section presents a summary of the different technologies used to detect emotions considering the various channels from which affective information can be obtained: emotion from speech, emotion from text, emotion from facial expressions, emotion from body gestures and movements, and emotion from physiological states [13].

**2.2.1. Emotion from Speech.** The voice is one of the channels used to gather emotional information from the user of a system. When a person starts talking, they generate information in two different channels: primary and secondary [14].

The primary channel is linked to the syntactic-semantic part of the locution (what the person is literally saying), while the secondary channel is linked to paralinguistic information of the speaker (tone, emotional state, and gestures). For example, someone says "That's so funny" (primary channel) with a serious tone (secondary channel). By looking at the information of the primary channel, the message received is that the speaker thinks that something is funny and by looking at the information received by the secondary channel, the real meaning of the message is worked out: the speaker is lying or being sarcastic.

Four technologies in this category can be highlighted: *Beyond Verbal* [15], *Vokaturi* [16], *EmoVoice* [17] and *Good Vibrations* [18]. Table 1 shows the results of the comparative study performed on the four analyzed technologies.

**2.2.2. Emotion from Facial Expressions.** As in the case of speech, facial expressions reflect the emotions that a person can be feeling. Eyebrows, lips, nose, mouth, and face muscles: they all reveal the emotions we are feeling. Even when a person tries to fake some emotion, still their own face is telling the truth. The technologies used in this field of emotion detection work in an analogous way to the ones used with speech: detecting a face, identifying the crucial points in the face which reveal the emotion expressed, and processing their positions to decide what emotion is being detected.

Some of the technologies used to detect emotions from facial expressions are *Emotion API* (Microsoft Cognitive Services) [19], *Affectiva* [20], *nViso* [21], and *Kairos* [22]. Table 2 shows a comparative study.

As far as the results are concerned, every tested technology showed considerable accuracy. However, several conditions (reflection on glasses and bad lightning) mask important facial gestures, generating wrong results. For

TABLE 1: Comparison of emotion detection technologies from speech.

Name	API/SDK	Requires Internet	Information returned	Difficulty of use	Free software
Beyond verbal	API	Yes	Temper, arousal, valence, and mood (up to 432 emotions)	Low	No
Votakuri	SDK	No	Happiness, neutrality, sadness, anger, and fear	Medium	Yes
EmoVoice	SDK	No	Determined by developer	High	Yes
Good vibrations	SDK	—	Happy level, relaxed level, angry level, scared level, and bored level	Medium	No

TABLE 2: Comparison of emotion detection technologies from facial expressions.

Name	API/SDK	Requires Internet	Information returned	Difficulty of use	Free software
Emotion API	API/SDK	Yes	Happiness, sadness, fear, anger, surprise, neutral, disgust, and contempt	Low	Yes (limited)
Affectiva	API/SDK	Yes	Joy, sadness, disgust, contempt, anger, fear, and surprise <sup>1</sup>	Low	Yes, with some restriction
nViso	API/SDK	No	Happiness, sadness, fear, anger, surprise, disgust, and neutral	—	No
Kairos	API/SDK	Yes	Anger, disgust, fear, joy, sadness, and surprise <sup>2</sup>	Low	Yes, only for personal use

<sup>1</sup>Besides, it also detects different facial expressions, gender, age, ethnicity, valence, and engagement. <sup>2</sup>Besides, it also detects user head position, gender, age, glasses, facial expressions, and eye tracking.

example, an expression of pain, in a situation in which eyes and/or brows cannot be seen, can be detected as a smile by these technologies (because of the stretching, open mouth).

As far as time is concerned, *Emotion API* and *Affectiva* show similar times to scan an image, while *Kairos* takes much longer to produce a result. Besides, the amount of values returned by *Affectiva* provides much more information to the developers, and it is easier to interpret the emotion that the user is showing than when we just have the weight of six emotions, for example. It is also remarkable the availability of *Affectiva*, which provides free services to those dedicated to research and education or producing less than \$1,000,000 yearly.

**2.2.3. Emotion from Text.** There are certain situations in which the communication between two people, or between a person and a machine, does not have the visual component inherent to face-to-face communication. In a world dominated by telecommunications, words are powerful allies to discover how a person may be feeling. Although emotion detection from text (also referred as sentiment analysis) must face more obstacles than the previous technologies (spelling errors, languages, and slang), it is another source of affective information to be considered. Since emotion detection from texts analyzes the words contained on a message, the process to analyze a text takes some more steps than the analysis of a face or a voice. There is still a model that needs to be trained, but now text must be processed in order to use it to train a model [23]. This processing involves tasks of tokenization, parsing and part-of-speech tagging, lemmatization, and stemming, among others. Four technologies of this category are *Tone Analyzer* [24], *Receptiviti* [25], *BiText* [26], and *Synesketch* [27].

Due to the big presence of social media and writing communication in the current society, this field is, along

with emotion detection from facial expressions, one of the most attractive fields to companies: posts from social media, messages sent to “Complaints” section, and so on. Companies which can know how their customers are feeling have an advantage over companies which cannot. Table 3 shows a comparative study of some of the key aspects of each technology. It is remarkable that as far as text is concerned, most of the companies offer a demo or trial version on their websites, while companies working on face or voice recognition are less transparent in this aspect. Regarding their accuracy, the four technologies have yielded good values. On the one hand, *BiText* has proved to be the simplest one, as it only informs if the emotion detected is good or bad. This way, the error threshold is wider and provides less wrong results. On the other hand, *Tone Analyzer* has proved to be less clear on its conclusions when the text does not contain some specific key words.

As far as the completeness of results is concerned, *Receptiviti* has been the one giving more information, revealing not only affective information but also personality-related information. The main drawback is that all these technologies (except *Synesketch*) are pay services and may not be accessible to everyone. Since *Synesketch* is not as powerful as the rest, it will require an extra effort to be used.

**2.2.4. Emotion from Body Gestures and Movement.** Even though people do not use body gestures and movement to communicate information in an active way, their body is constantly conveying affective information: tapping with the foot, crossing the arms, tilting the head, changing our position a lot of times while seated, and so on. Body language reveals what a person is feeling in the same way our voice does.

However, this field is quite new, and there is not a clear understanding about how to create systems able to detect emotions relating to body language. Most researchers have

TABLE 3: Comparison of emotion detection technologies from text.

Name	API/SDK	Requires Internet	Information returned	Difficulty of use	Free software
Tone analyzer	API	Yes	Emotional, social, and language tone	Low	No
Receptiviti	API	Yes	See [29]	Low	No
BiText	API	Yes	Valence (positive/negative)	Low	No
Synesketch	SDK	No	Six basic emotions	Medium	Yes

focused on facial expressions (over 95 per cent of the studies carried out on emotions detection have used faces as stimuli), almost ignoring the rest of channels through which people reveal affective information [28].

Despite the newness of this field, there are several proposals focused on recognizing emotional states from body gestures, and these results are used for other purposes. Experimental psychology has already demonstrated how certain types of movements are related to specific emotions [29]. For example, people experimenting fear will turn their bodies away from the point which is causing that feeling; people experimenting happiness, surprise, or anger will turn their bodies towards the point causing that feeling.

Since there are no technologies available for emotion detection from body gestures, there is not any consensus about the data we need to detect emotions in this way. Usually, experiments on this kind of emotion detection use frameworks (as for instance, SSI) or technologies to detect the body of the user (as for instance, Kinect), so the researches are responsible for elaborating their own models and schemes for the emotion detection. These models are usually built around the joints of the body (hands, knees, neck, head, elbows, and so on) and the angle between the body parts that they interconnect [30], but in the end, it is up to the researchers.

**2.2.5. Emotion from Physiological States.** Physiologically speaking, emotions originate on the limbic system. Within this system, the amygdala generates emotional impulses which create the physiological reactions associated with emotions: electric activity on face muscles, electrodermal activity (also called galvanic skin response), pupil dilatation, breath and heart rate, blood pressure, brain electric activity, and so on. Emotions leave a trace on the body, and this can be measured with the right tools.

Nevertheless, the information coming directly from the body is harder to classify, at least with the category system used in other emotion detection technologies. When working with physiological signals, the best option is to adopt a classification system based on a dimensional approach [25]. An emotion is not just “happiness” or “sadness” anymore, but a state determined by various dimensions, like valence and arousal. It is because of this that the use of physiological signals is usually reserved for research and studies, for example, related to autism. There are no emotion detection services available for this kind of detection based on physiological states, although there are plenty of sensors to read these signals.

In a recent survey on mobile affective computing [31], authors make a thorough review of the current literature on

affect recognition through smartphone modalities and show the current research trends towards mobile affective computing. Indeed, the special capacities of mobile devices open new research challenges in the field of affective computing that we aim to address in the mobile version of the system proposed.

Finally, we can also find available libraries to be used in different IDEs (integrated development environments) supporting different programming languages. For instance, NLTK, in python [32] can be used to analyze natural language for sentiment analysis. Scikit-learn [33], also in python, provides efficient tools for data mining and data analysis with machine learning techniques. Lastly, OpenCV (Open Source Computer Vision Library) [34] supports C++, Python, and Java interfaces in most operating systems. It is designed for computer vision and allows the detection of elements caught by the camera in real time to analyze the facial points detected according, for instance, to the Facial Action Coding System (FACS) proposed by Ekman and Rosenberg [35]. The data gathered could be subsequently processed with the scikit-learn tool.

```

aff_information = get_affective_information()
#aff_information = {"face": [...], "voice": [...],
"mimic": [...]}
stress_flags = {"face": 0.0, "voice": 0.0, "mimic": 0.0}
#values from 0 to 1 indicating stress levels detected
for er_channel, measures in aff_information:
    measure_stress(er_channel, measures, stress_flags)
if (stress_flags["face"] > 0.6 and
    stress_flags["voice"] < 0.3 and
    stress_flags["mimic"] < 0.1):
    #reaction to affective state A
if (stress_flags["face"] < 0.1 and
    stress_flags["voice"] < 0.1 and
    stress_flags["mimic"] < 0.5):
    #reaction to affective state B
...

```

### 3. Modifying the Behaviour of an Educational Software Application Based on Emotion Recognition

Human interaction is, by definition, multimodal [36]. Unless the communication is done through phone or text, people can see the face of the people they are talking to, listen to their voices, see their body, and so on. Humans are, at this



point, the best emotion detectors as we combine information from several channels to estimate a result. This is how multimodal systems work.

It is important to remark that a multimodal system is not just a system which takes, for example, affective information from the face and from the voice and calculates the average of each value. The hard part of implementing one of these systems is to combine the affective information correctly. For example, a multimodal system combining text and facial expressions that detects a serious face and the message “it is very funny” will return “sarcasm/lack of interest,” while the result of combining these results in an incorrect way will return “happy/neutral.” It is proven that by combining information from several channels, the accuracy of the classification improves significantly.

For example, let us imagine we need to assess the stress levels of a person considering the affective information gathered through three different channels: affective information extracted from facial expressions, voice, and body language. Since we have more than one channel, we can support each measure taken from each channel with values detected in the others.

This way, it is possible not only to confirm with a high level of certainty the occurrence of an affective state, but also to detect situations that could not be sensed without performing multimodal emotion detection, as sarcasm.

The following code snippet shows an easy example of affective information combination. The mere fact of considering a measure in the context of more affective information gives us a whole new dimension of information.

To this end, we have developed an initial prototype in order to study how using multimodal emotion detection systems on educational software applications could enhance the user experience and performance. The proposed prototype, named emoCook, has been developed as a game to teach English to 9–11-year-old children. Information about this prototype can be found at [37]. At present, the prototype is only available in Spanish as it is initially addressed to Spanish-speaking children in the process of learning English.

The architecture of this application is shown in Figure 1. During the gameplay, the user is transmitting affective information (Figure 1-1) through their face, their voice, their behaviour, and so on. The prototype is receiving this information (Figure 1-3) and sending it to several third-party emotion detection services (Figure 1-4). After retrieving this information (Figure 1-5), we put it in context to extract conclusions from it about the user’s performance (Figure 1-6). Based on these results, the pace and difficulty level of the game changes (Figure 1-7), adapting it to the user’s affective state (Figure 1-2).

The theme of the game was focused on cooking issues to practice vocabulary and expressions related to this topic. It is organized in different recipes, from the easiest to hardest. Each recipe is an independent level and is divided into two parts. The first part is a platform game in which the player must gather all the ingredients needed to cook the recipe (Figure 2). The ingredients are falling from the sky all the time, along with other food we do not need for the recipe. If the player catches any food that is not in the ingredients list,

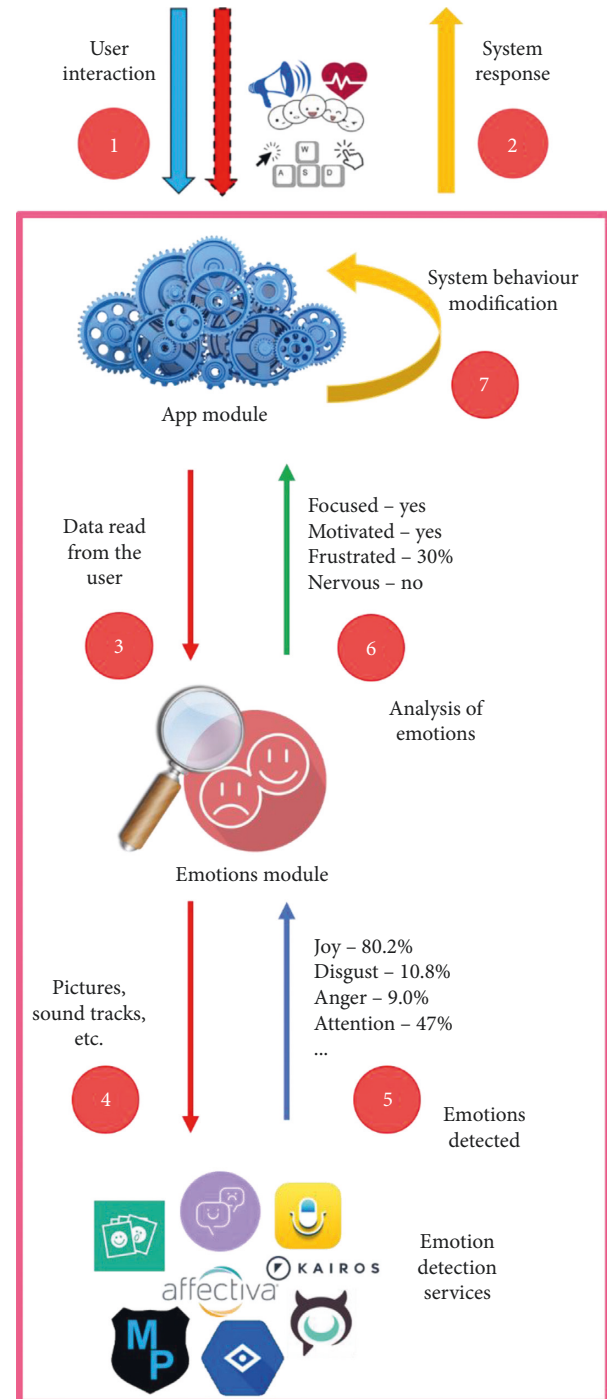


FIGURE 1: Application architecture.

it is considered as a mistake. The maximum number of mistakes allowed per level is five.

After finishing this first part, the system shows a set of sentences (more or less complex) including vocabulary related to the recipe that the player has to read out loud to practice speaking and pronunciation. If the user fails thrice to read a sentence, the system will move to the next one, or finish the exercise if it is the last sentence.

This prototype has been implemented with three emotion detection technologies, which monitor the player’s



FIGURE 2: emoCook prototype.

affective state, and the results obtained are used to change the difficulty level and the pace of the game. Each time the player finishes a level, the affective data are analyzed, and according to the results, the difficulty of the next level is set. The technologies integrated in the system are the following:

- (i) *Afectiva*. It uses the camera feed to read the facial expression of the player.
- (ii) *Beyond Verbal*. It gathers the audio collected during the speech exercise to identify the affective state of the player attending to their speech features.
- (iii) *Keylogger*. The game keeps a record of the keys pressed by the users, considering different factors: when they press a correct key, when they do not, when they press it too fast, and so on.

Because of changes on Beyond Verbal API, affective data from the speech could not be collected, so in the end, only data from the facial expression (using Affectiva) and from the behaviour when pressing keys (using Keylogger) were used. Affectiva is a third-party service, while Keylogger was developed within the prototype.

A mobile version of the system is also available, and it can be used through a browser running on a mobile device [37]. This way, the game can be controlled both with the arrow keys in a keyboard and by touching on a tactile screen. Touching on the left-hand side of the screen makes the character move to the left. Touching on the right-hand side of the screen makes the character move to the right and touching twice very quickly in any part of the screen makes the character jump upwards.

Figure 3 shows a screenshot of the mobile version of the application running in the Firefox browser in a mobile device. The possibility of using the system through a mobile device opens new ways of detecting emotions that we aim to explore in further research. For instance, we could use sensors such as the accelerometer or gyroscope to gather affective information. Initial trials have been performed with



FIGURE 3: Mobile version of emoCook system.

the API offered in [38] with promising results that will be further explored.

## 4. Evaluation of the System

In order to prove the initial hypothesis, the system has been assessed with real users by applying the method described in this section.

**4.1. Participants and Context.** We recruited sixteen children aged between 10 and 11 years old belonging to the same primary school and with a similar level of English knowledge to avoid differences in the education level that could affect the evaluation results. Their parents had been previously informed and authorised their participation in this evaluation. The setup of the experiment consisted of two laptops, one in front of the other so that participants could not see each other. Both laptops were equipped with mouse and webcam and Windows 10 as operating system and were connected to the same Wi-Fi network. The prototype was accessed through the browser Google Chrome in both laptops. This setup was prepared in a room the English

teachers of the primary school provided us within the school premises.

**4.2. Evaluation Metrics.** The system was measured considering three types of metrics: effectiveness, efficiency, and satisfaction, that is, the users' subjective reactions when using the system. Effectiveness was measured by considering task completion percentage, error frequency, and frequency of assistance offered to the child. Efficiency was measured by calculating the time needed to complete an activity, specifically, the mean time taken to achieve the activity. Besides, some other aspects were also considered such as the number of attempts needed to successfully complete a level, number of keystrokes, and the number of times a key was pressed too fast as an indicative signal of nervousness.

Finally, satisfaction was measured with the System Usability Scale (SUS) slightly adapted for teenagers and kids [39]. This questionnaire is composed of ten items related to the system usage. The users had to indicate the degree of agreement or disagreement on a 5-point scale.

**4.3. Experimental Design.** After several considerations regarding the evaluation process for games used in learning environments [40], the following features were established:

- (i) *Research Design.* The sample of participants was divided into two groups of the same size, being one of them the control group. This control group tested the application implemented without emotion detection and hence without modifying the behaviour of the application in real time according to the child's emotions. This one was called the System 2 group. The other group tested the prototype implemented with emotion detection which adapted its behaviour, by modifying the pace of the game and difficulty level, according to the emotions detected on the user, in such a way that if the user becomes bored, the system increases the pace of the game and difficulty level and on the contrary, if the user becomes stressed or nervous, the system decreases the speed of the game and difficulty level. This one was called the System 1 group. By doing this, it can be shown how using emotion detection to dynamically vary the difficulty level of an educational software application influences the performance and user experience of the students.
- (ii) *Intervention.* The test was conducted in the premises of the primary school in a quiet room where just the participants (two at a time) using System 1 and System 2 and the evaluators were present. We prepared two laptops of similar characteristics, one of them running System 1 with the version of the application implemented with emotion recognition and the other laptop running System 2 with the version of the application without emotion recognition.

The whole evaluation process was divided into two parts:

- (i) *Introduction to the Test.* At the beginning of the evaluation, the procedure was explained to the sixteen children at a time, and the game instructions for the different levels were given.
- (ii) *Performing the Test.* Kids were called in pairs to the room where the laptops running System 1 and System 2 were prepared. None of the children knew what system they were going to play with. At the end of the evaluation sessions, the sixteen children completed the SUS questionnaire. Researchers were present all the time, ready to assist the participants and clarify doubts when necessary. When a participant finished the test, they returned to their classroom and called the next child to go in the evaluation room.

To keep the results of each participant fully independent, the sixteen users were introduced on the database of the prototype with the key "evalX," being "X" a number. Users with an odd "X" used System 1, while those with an even "X" were assigned to System 2 (control group).

The task that the participants had to perform was to play the seven levels of the prototype, including each level a platform game and a reading out loud exercise. The data collected during the evaluation sessions were subsequently analysed, and the outcomes are described next.

**4.4. Evaluation Outcomes and Discussion.** Although participants with System 1 needed, on average, a bit more time per level to finish (76.18 seconds against 72.7), we could appreciate an improvement on the performance of the participants using System 1, as most of them made less than 5 mistakes on the last level, while only one of the control group users of System 2 had less than 5 mistakes.

Figure 4 shows the evolution of the average number of mistakes, which increases in the control group (System 2) from level 4 onwards. Since the game adapts its difficulty (in System 1), after detecting a peak of mistakes in the fourth level (as a sign of stress, detected as a combination of negative feelings found in the facial expression and the way the participant used the keyboard), the difficulty level was reduced. This adaptation made the next levels easier to play for participants using System 1, what was reflected in less mental effort. Since participants using System 2 did not have this feature, their average performance got worse.

On average, participants using System 1 needed 1.33 attempts to finish each level, while participants using System 2 needed 1.59, almost 60% more. Also, the ratio of mistakes to total keystrokes was also higher in the case of System 2 users (19% against the 12% from users of System 1). Likewise, System 2 users asked for help more often (13 times) than System 1 users (10 times). In future experimental activities, the sample size would be increased in order to obtain more valuable data.

The evaluation was carried out as a between-subjects design with *emotion recognition* as the independent variable (using or not using emotion recognition features) and

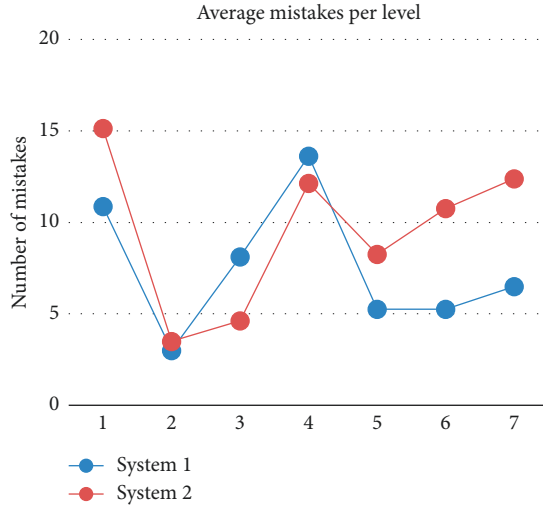


FIGURE 4: Average mistakes per level.

*attempts* (attempts needed to finish each level), *time* (time (seconds) needed to finish each level), *mistakes* (number of mistakes), *keystrokes* (number of keystrokes), and *stress* (number of times a key was pressed too fast in a short time) as the dependent variables.

We performed a standard  $t$ -test [41] to compare the means of each dataset and test the null hypothesis that there was no significant difference in the students' performance when using emotion recognition to adapt the system behaviour. We used  $\alpha = 0.05$  as our limit for statistical significance, with significant results reported below.

Regarding *keystrokes* ( $t = 0.97$ ;  $p = 0.666$ ), *mistakes* ( $t = -1.51$ ;  $p = 0.26$ ), and *stress* ( $t = 1.13$ ;  $p = 0.51$ ),  $t$ -test results confirmed the null hypothesis was false and, thus, that the two datasets are significantly different.

Although the dependent variables *time* ( $t = 0.44$ ;  $t = 1.31$ ) and *attempts* ( $t = -0.42$ ;  $t = 1.33$ ) were similar in both datasets, the efficiency (considered as the lowest number of actions a user needs to finish each level) is greater in users of System 1, even though both users of System 1 and System 2 finished within a similar time frame, what helped the first ones to make less mistakes. The outcomes of the evaluation shown in Figure 4 indicate a clear improvement when using System 1 as the number of mistakes increases in users of System 2 at higher difficulty levels.

Finally, Table 4 and Table 5 show the results of the SUS scores per system and participant. The final value is between 0 and 100, 100 being the highest degree of user's satisfaction. As we can see, System 1 users rated the application with a higher level of satisfaction compared to the level obtained by users of System 2, as shown in Figure 5.

## 5. Conclusions and Final Remarks

Emotion detection, together with Affective Computing, is a thriving research field. Few years ago, this discipline did not even exist, and now there are hundreds of companies working exclusively on it, and researchers are investing time and resources on building affective applications. However,

TABLE 4: Satisfaction results for System 1.

Participant	SUS score
1	90
3	90
5	90
7	100
9	92.5
11	87.5
13	82.5
15	80
Mean	89.06

TABLE 5: Satisfaction results for System 2.

Participant	SUS score
2	92.5
4	75
6	87.5
8	85
10	75
12	92.5
14	72.5
16	90
Mean	83.75

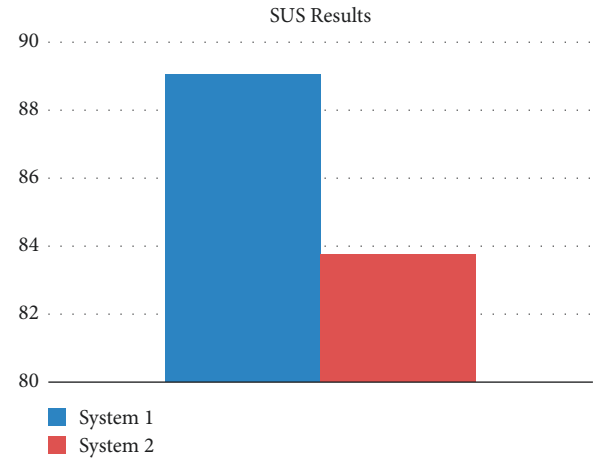


FIGURE 5: Comparison of SUS results in both systems.

emotion detection has still many aspects to improve in the coming years.

Applications which obtain information from the voice need to be able to work in noisy environments, to detect subtle changes, maybe even to recognize words and more complex aspects of human speech, like sarcasm.

The same applies for applications that detect information from the face. Most people use glasses nowadays, which can greatly complicate accurate detection of facial expressions.

Applications able to read body gestures do not even exist now, even though it is a source of affective information as valid as the face. There are already applications for body detection (Kinect), but there is no technology like *Affectiva* or *Beyond Verbal* for the body yet.

Physiological signals are even less developed, because of the imposition of sensors that this kind of detection requires.



However, some researchers are working on this issue so physiological signals can be used as the face or the voice. In a not too distant future, reading the heartbeat of a person with just a mobile with Bluetooth may not be as crazy as it may sound.

Previous technologies analyze the impact of an emotion in our bodies, but what about our behaviour? A stressed person usually tends to make more mistakes. In the case of a person interacting with a system, this will be translated in faster movements through the user interface, or more mistakes when selecting elements or typing, and so on. This can be logged and used as another indicator of the affective state of a person.

All these technologies are not perfect. Humans can see each other and estimate how other people are feeling within milliseconds, and with a small threshold error, but these technologies can only try to figure out how a person is feeling according to some input data. To get more accurate results, more than one input is required, so multimodal systems are the best way to guarantee results with the highest levels of accuracy.

In this paper, we present an educational software application that incorporates affective computing by detecting the users' emotional states to adapt its behaviour to the emotions detected. Assessing this application in comparison with another version without emotion detection, we can conclude that the user experience and performance is higher when including a multimodal emotion detection system. Since the system is continuously adapting itself to the user according to the emotions detected, the level of difficulty adjusts much better to their real needs.

On the basis of the outcomes of this research, new challenges and possibilities in other kind of applications will be explored; for example, we could "stress" a user in a game if the emotions detected show that the user is bored. The application could even introduce dynamically other elements to engage the user in the game. What is too simple bores a user, whereas what is too complex causes anxiety. Changing the behaviour of an application dynamically according to the user's emotions, and also according to the nature of the application, increases the satisfaction of the user and helps them decrease the number of mistakes.

As future work, among other things, we aim to improve the mobile aspects of the system and explore further the challenges that the sensors offered by mobile devices bring about regarding emotion recognition, especially in educational settings.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This research work has been partially funded by the regional project of JCCM with reference SBPLY/17/180501/000495, by

the Scholarship Program granted by the Spanish Ministry of Education, Culture and Sport, and the predoctoral fellowship with reference 2017-BCL-6528, granted by the University of Castilla-La Mancha. We would also like to thank the teachers and pupils from the primary school "Escolapios" who collaborated in the assessment of the system.

## References

- [1] R. W. Picard, *Affective Computing*, MIT Press, Cambridge, UK, 1997.
- [2] E. Johnson, R. Hervás, C. Gutiérrez, T. Mondéjar, and J. Bravo, "Analyzing and predicting empathy in neurotypical and nonneurotypical users with an affective avatar," *Mobile Information Systems*, vol. 2017, Article ID 7932529, 11 pages, 2017.
- [3] S. Koelstra, C. Muhl, M. Soleymani et al., "DEAP: a database for emotion analysis using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012.
- [4] R. Kaliouby, "We need computers with empathy," *Technology Review*, vol. 120, no. 6, p. 8, 2017.
- [5] S. L. Marie-Sainte, M. S. Alrazgan, F. Bousbahi, S. Ghoulali, and A. W. Abdul, "From mobile to wearable system: a wearable RFID system to enhance teaching and learning conditions," *Mobile Information Systems*, vol. 2016, Article ID 8364909, 10 pages, 2016.
- [6] M. Li, Y. Xiang, B. Zhang, and Z. Huang, "A sentiment delivering estimate scheme based on trust chain in mobile social network," *Mobile Information Systems*, vol. 2015, Article ID 745095, 20 pages, 2015.
- [7] B. Ovcjak, M. Hericko, and G. Polancic, "How do emotions impact mobile services acceptance? A systematic literature review," *Mobile Information Systems*, vol. 2016, Article ID 8253036, 18 pages, 2016.
- [8] P. Williams, "Emotions and consumer behavior," *Journal of Consumer Research*, vol. 40, no. 5, pp. viii–xi, 2014.
- [9] E. Andrade and D. Ariely, "The enduring impact of transient emotions on decision making," *Organizational Behavior and Human Decision Processes*, vol. 109, no. 1, pp. 1–8, 2009.
- [10] R. W. Picard, "Affective computing: challenges," *International Journal of Human-Computer Studies*, vol. 59, no. 1–2, pp. 55–64, 2003.
- [11] R. W. Picard, "Affective computing," Tech. Rep. 321, M.I.T Media Laboratory Perceptual, Computing Section, Cambridge, UK, 1995.
- [12] I. Morgun, *Types of Machine Learning Algorithms*, 2015.
- [13] J. García-García, V. Penichet, and M. Lozano, "Emotion detection: a technology review," in *Proceedings of XVIII International Conference on Human Computer Interaction*, Cancún, México, September 2017.
- [14] S. Casale, A. Russo, G. Scebbba, and S. Serrano, "Speech emotion classification using machine learning algorithms," in *Proceedings of IEEE International Conference on Semantic Computing 2008*, pp. 158–165, Santa Monica, CA, USA, August 2008.
- [15] Beyond Verbal, "Beyond verbal—the emotions analytics," May 2017, <http://www.beyondverbal.com/>.
- [16] Vokaturi, May 2017, <https://vokaturi.com/>.
- [17] T. Vogt, E. André, and N. Bee, "EmoVoice—a framework for online recognition of emotions from voice," in *Perception in Multimodal Dialogue Systems*, E. André, L. Dybkjær, W. Minker, H. Neumann, R. Pieraccini, and M. Weber, Eds., Springer, Berlin, Heidelberg, Germany, 2008.

- [18] Good Vibrations, “Good vibrations company B.V.–recognize emotions directly from the voice,” May 2017, <http://good-vibrations.nl>.
- [19] Microsoft, “Microsoft cognitive services–emotion API,” May 2017, <https://www.microsoft.com/cognitive-services/en-us/emotion-api>.
- [20] Affectiva, “Affectiva,” May 2017, <http://www.affectiva.com>.
- [21] nViso, “Artificial intelligence emotion recognition software–nViso,” May 2017, <http://nviso.ch/>.
- [22] Kairos, “Face recognition, emotion analysis & demographics,” May 2017, <https://www.kairos.com/>.
- [23] H. Binali and V. Potdar, “Emotion detection state of the art,” in *Proceedings of the CUBE International Information Technology Conference on-CUBE '12*, pp. 501–507, New York, NY, USA, September 2012.
- [24] IBM, “Tone analyzer,” May 2017, <https://tone-analyzer-demo.mybluemix.net/>.
- [25] Receptiviti, May 2017, <http://www.receptiviti.ai/>.
- [26] Bitext, “Bitext API,” May 2017, <https://api.bitext.com>.
- [27] U. Krčadinac, “Synesketch: free open-source textual emotion recognition and visualization,” May 2017, <http://krčadinac.com/synesketch/>.
- [28] A. Kleinsmith and N. Bianchi-Berthouze, “Affective body expression perception and recognition: a survey,” *IEEE Transactions on Affective Computing*, vol. 4, no. 1, pp. 15–33, 2013.
- [29] R. W. Picard, “Future affective technology for autism and emotion communication,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1535, pp. 3575–3584, 2009.
- [30] Universität Augsburg University, “OpenSSI,” May 2017, <https://hcm-lab.de/projects/ssi/>.
- [31] E. Politou, E. Alepis, and C. Patsakis, “A survey on affective computing,” *Computer Science Review*, vol. 25, pp. 79–100, 2017.
- [32] N. Hardeniya, *NLTK Essentials*, Packt Publishing Limited, Birmingham, UK, 2015.
- [33] G. Hackeling, *Mastering Machine Learning with Scikit-Learn*, Packt Publishing Limited, Birmingham, UK, 2014.
- [34] J. Howse, P. Joshi, and M. Beyeler, *OpenCV: Computer Vision Projects with Python*, Packt Publisher Limited, Birmingham, UK, 2016.
- [35] P. Ekman and E. Rosenberg, *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*, Oxford University Press, Oxford, UK, 2005.
- [36] J. Tao and T. Tan, “Affective computing: a review,” *Lecture Notes in Computer Science*, vol. 3784, pp. 981–995, 2005.
- [37] J. M. Garcia-Garcia, “emoCook,” December 2017, <https://emocook.herokuapp.com/>.
- [38] A. Bar, “What web can do today. An overview of the device integration HTML5 APIs,” June 2018, <https://whatwebcando.today/device-motion.html>.
- [39] Usability.gov, “Usability.gov–improving the user experience,” January 2018, <https://www.usability.gov/get-involved/blog/2015/02/working-with-kids-and-teens.html>.
- [40] A. All, E. P. Nuñez Castellar, and J. Van Looy, “Assessing the effectiveness of digital game-based learning: best practices,” *Computers & Education*, vol. 92–93, pp. 90–103, 2016.
- [41] D. Garson, *Significance Testing: Parametric and Nonparametric*, Statistical Associates Publishing, Blue Book Series, Asheboro, NC, USA, 2012.

